UWM Data Publishing Guide

Researchers with NSF, NIH[1], DOE, and other federal funding are now expected to make data from their research publicly available. Additionally, journals such as PLOS, Science, Nature, BMJ, and more require researchers to make their data available after publication and/or during peer review. This guide is intended to help UWM researchers navigate these new requirements for data sharing.

If you have further questions about data management and sharing, contact data-services@uwm.edu. You can also find more information at:

UWM Data Services: http://uwm.edu/libraries/dataservices/
Data management guide: http://guides.library.uwm.edu/data

When to share

Data sharing happens once you publish. After you publish an article, white paper, etc. you should make the data that support that research publicly available.

What to share

Share data that support your publications and anything necessary to reproduce your results. The form of data to share (raw, cleaned up, analyzed, etc.) varies by field and often depends other considerations, like file size. Use your best judgment as to what is the most useful form for your data.

Sensitive data, such as that containing personally identifiable information, should not be shared. However, do consider sharing if you can properly anonymize or redact the data. Corporate-sponsored research may also have similar restrictions on sharing; contact Mark Doremus (doremus@uwm.edu) if you have private funding and want to share your data.

Why to share

Data sharing has been shown to increase citations for the corresponding article (see https://peerj.com/articles/175/). Data sharing also makes it easier for you to find and reuse data later as the data live in a stable location. Finally, all major federal funding agencies will soon start requiring data sharing if they do not already (see http://1.usa.gov/1jg1QXt).

Preparing to share

Perform quality control on your data prior to sharing to clean up errors and inconsistencies. OpenRefine (http://openrefine.org/) is particularly recommended for cleaning up tabular data.

---

[1] NIH grants for over $500,000 per year are required to make the data publicly accessible

Use non-proprietary or common file types (.csv, .txt, etc.) whenever possible. This makes data files easier to open and more long lived.

Shared data should also be documented. Use a README.txt file, a data dictionary, or some other form of documentation so that others can understand your data. Data Services' data management guide has more information on documentation formats: http://guides.library.uwm.edu/data.

It is also recommended to license your data with a Creative Commons license (http://creativecommons.org/), preferably CC0 as you cannot copyright most research data. Copyright may apply to some types of data, such as images, but the majority of data exist outside of copyright. An open license clears up any confusion over copyright and what rights are available for the data.

For more recommendations, see the article "*Nine simple ways to make it easier to (re)use your data*": http://library.queensu.ca/ojs/index.php/IEE/article/view/4608

Where to share

Share your data in a data repository. This is preferable to sharing on a personal website or by request because repositories are more stable and this method of sharing requires no work from you after deposit. Additionally, many repositories have built-in methods for citing your data.

To find a repository for your data, consider using one of the following:

- figshare (general): http://figshare.com/
- Dryad (biology): http://datadryad.org/
- ICPSR (social science): https://www.icpsr.umich.edu
- UWM Digital Commons (small, discrete datasets): http://dc.uwm.edu/

There are many other general and disciplinary repositories available. Contact Data Services at data-services@uwm.edu for a recommendation.

Other considerations

Consider sharing your research code via Google Code (https://code.google.com/) or GitHub (https://github.com/), which helps with reproducibility.

Add datasets to your CV and report them as products of your grant – this is especially recommended when applying for your next grant. You can also sometimes track metrics (views, downloads, etc.) via the data repository or ImpactStory (https://impactstory.org/).